MULTISCALE MODEL. SIMUL.
Vol. 5, No. 4, pp. 1248–1263

LOW-RESOLUTION MOLECULAR DYNAMICS SIMULATIONS OF THE 30S RIBOSOMAL SUBUNIT*

QIZHI CUI†, ROBERT K.-Z. TAN‡, STEPHEN C. HARVEY‡, AND DAVID A. CASE†

Abstract. Low-resolution molecular models can provide appropriate and efficient ways for studying large biomolecular systems such as the ribosome. We have developed computer codes that use the Yammp Under Python modeling package to assemble low-resolution force fields for RNA-protein complexes, and that connect these to the Amber molecular simulation package. This pipeline combines many of the complementary strengths of these two packages. Our target here is the 30S ribosomal subunit from *Thermus thermophilus*. One hundred nanosecond Langevin dynamics simulations were performed for the bound and the unbound 16S RNA, and conformational changes of the 16S RNA and its interaction with the 30S proteins were examined to establish the fidelity of our model. The S7 protein assembly pathway was also examined, and the effects of protein binding order on the 16S RNA were analyzed. The simulations suggest that ribosomal proteins play important roles in maintaining the native 16S RNA structure. "Primary" proteins (in terms of assembly) help more in stabilizing the conformation of the RNA than do secondary and tertiary proteins. Ribosomal proteins appear to bind to the RNA in an organized fashion wherein primary and secondary proteins help to prepare the binding sites for tertiary proteins. The methodology and tools described here should provide useful ways to explore other aspects of ribosomal conformational changes by means of molecular dynamics simulations.

Key words. multiscale, ribosome, molecular, dynamics

AMS subject classifications. 82C31, 92E10, 92C40

DOI. 10.1137/05064850X

1. Introduction.

Modeling of the ribosome can be traced back to the 1980s when three-dimensional models of the small subunit of *Escherichia coli* were built in several labs [1, 2, 3, 4, 5, 6, 7]. The modeling was done manually and tediously and was based on the available secondary structures and some very low-resolution structural data obtained from phylogenetic studies, cross-linking and footprinting experiments, chemical accessibility, electron microscopy, mutational studies, and so on. In the 1990s, methods based on molecular mechanics were used to refine three-dimensional structures of the small ribosomal subunit based on the accumulated structural data [8, 9, 10].

Starting in 2000, atomic resolution crystal structures of small and large subunits and the whole ribosome were determined [11, 12, 13, 14, 15, 16]. This high-resolution structural data provides great opportunities for theoreticians to examine structural, dynamic, and electrostatic properties of the ribosomes with modern computer modeling and simulation techniques. For example, by means of rigid body Monte Carlo simulations, Stagg, Mears, and Harvey [17] built a structural model at a molecular level for the 30S subunit from *Thermus thermophilus* and examined the S7 pathway of protein assembly. Li, Ma, and Shapiro [18] studied binding interactions between

*Received by the editors December 27, 2005; accepted for publication (in revised form) March 16, 2006; published electronically December 28, 2006.

http://www.siam.org/journals/mms/5-4/64850.html

†Department of Molecular Biology, TPC15, The Scripps Research Institute, La Jolla, CA 92037 (qzcui@scripps.edu, case@scripps.edu). The fourth author's research was supported by an NIH grant (RR12255).

‡School of Biology, 310 Ferst Drive, Georgia Institute of Technology, Atlanta, GA 30332-0230 (rt106@mail.gatech.edu, stephen.harvey@biology.gatech.edu). The third author's research was supported by an NIH grant (RR12255).

the core central domain of the 16S RNA and the ribosomal protein S15 by MD simulations. Two groups have used a coarse-grained elastic network model to study the ratchet-like rearrangements of the 70S ribosome [19, 20]. Trylska, Tozzini, and McCammon [21] reported results of molecular dynamics (MD) simulations of the 70S ribosome from *Thermus thermophilus* using DL_POLY programs [22] with a coarse-grained model for the ribosome. In an even larger simulation, Sanbonmatsu, Joseph, and Tung [23] have computed totally about 20 nanosecond targeted MD simulation using an all-atom model.

However, it is still very challenging to apply MD simulation methodologies to large complexes such as the ribosome, due to their tremendous size and long time scales of conformational change. For example, the simulations of Sanbonmatsu, Joseph, and Tung required about 1 million CPU hours on the ASCI LANL Qmachine, a Hewlett Packard supercomputer with Alpha EV processors and Elan interconnect. Coarse-grained models are justified in part by the need to reduce computational burden and in part by the fact that one does not always need all-atom detail to study these large systems.

In this paper, we tackle two tasks. One is the implementation of a suitable coarse-grained model for complexes containing both proteins and nucleic acids. This follows techniques developed by Tan and Harvey [24], who built low-resolution three-dimensional models for large RNAs and ribonucleoprotein particles [8, 9, 10, 25] using the Yammp package. The second task is to make this modeling approach available to a wider user community by connecting the Yammp implementation of this model to Amber. This takes advantage of Amber's wide distribution and of the fact that Amber has been parallelized, where Yammp has not.

Recently, Yammp has been reorganized as an extension of Python; the new version of the program package is called Yammp Under Python (YUP) [26, 27]. In this work, we employ representations and force fields as [25] but with some updated potential parameters. Using an early version of YUP, we completed force field assembly (FFA) and the related modules for our specific low-resolution models for the ribosomal RNA and proteins. Although it is possible to carry out MD simulations in the YUP package, there are some significant advantages in using a more conventional program package such as Amber for this task. First, Amber is a parallel code that is tuned for efficiency on many machines, facilitating large-scale simulations. Second, Amber has a well-developed suite of trajectory analysis programs, and many visualization programs recognize its file formats. In order to combine the best features of YUP and Amber, a Python script was developed to integrate reading in PDB files for an initial conformation, adding coordinates for space-filling pseudoatoms, running YUP modules for assembling force fields, and outputting Amber parameter and coordination files for MD simulations. Here we report our initial applications of this methodology to study the 30S subunit from *Thermus thermophilus* and analyze the conformational changes of the 16S RNA and roles of the 30S proteins. Several 100 nanosecond Langevin dynamics simulations were performed for the bound and the unbound 16S RNA. The protein binding events and the 16S RNA conformational changes were analyzed from a dynamic point of view.

2. Models and methods.

2.1. Low-resolution model for the 30S subunit. Our low-resolution model for the 16S RNA was constructed based on the secondary structure of the 16S RNA from the small subunit of the *Thermus thermophilus* [15, 28, 29] (see Figure 1(a)). The model follows the design of earlier Yammp coarse-grained RNA models [25]. Each

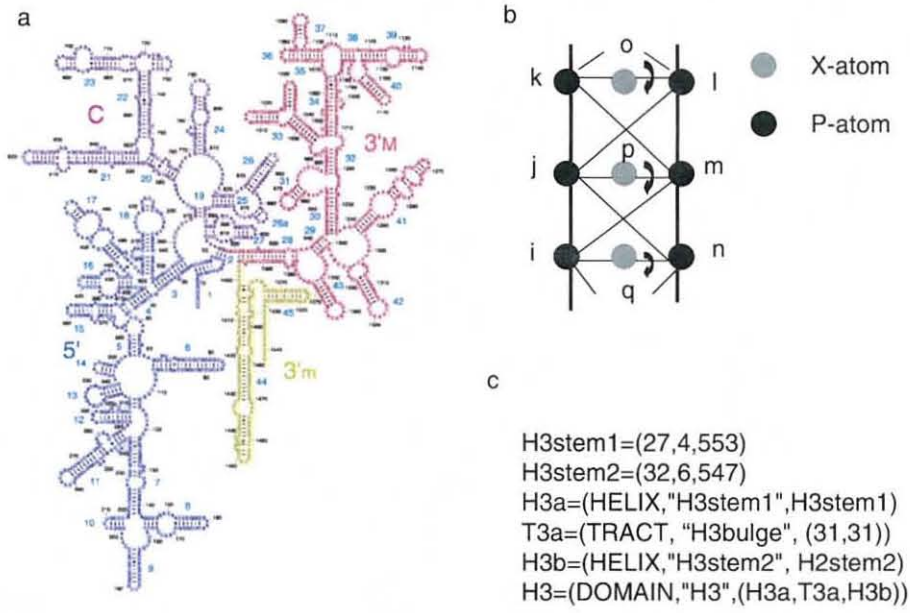


FIG. 1. (a) *Secondary structure of the 16S RNA from Thermus thermophilus with the 5′, central (C), 3′ major (3′M), and 3′ minor (3′m) domains shaded in blue, purple, red, and green, respectively. This figure is from* [15] *and is reprinted with permission.* ©2001, *AAAS. The original schematic representation is from* [28, 29]. (b) *Model of a helix stem, including P atoms (i, j, k, l, m, n) and X-atoms (o, p, q), along with the pseudobonds required to define the three-dimensional geometry of a double helix. The lengths of these pseudobonds and the accompanying angles are chosen to match those of an ideal A-RNA double helix. The pseudotorsions (e.g., i-j-m-n) are required to guarantee that the helix is right-handed, since the mirror image of such a helix, though left-handed, would satisfy the bond and angle restraints.* (c) *Definition of helix H3 in the* `ReadyMade.py` *module.*

nucleotide is represented by a single pseudoatom (P-atom) centered on the phosphate atom. The region which consists of three or more contiguous base-paired nucleotides is considered to be a helix stem (see Figure 1(b)). For each helix stem, space-filling pseudoatoms (X-atoms) are placed at the geometric center of each base pair but not at the end base pairs. The X-atoms are used to prevent helix interpenetration.

Harmonic potentials are used for bonds, angles, and torsions.

Bond:

$$K_b(r_{ij} - b_o)^2, \tag{1}$$

angle:

$$K_a(\theta_{ijk} - \theta_o)^2, \tag{2}$$

torsion:

$$K_t(\phi_{ijmn} - \phi_o)^2, \tag{3}$$

where i, j, k, m, n denote sites of pseudoatoms; K_b, K_a, K_t are the force constants; r_{ij}, θ_{ijk}, ϕ_{ijmn} are the instantaneous bond length, angle, and torsion values; b_o, θ_o, ϕ_o are the corresponding equilibrium or ideal values. The potential parameters are

TABLE 1

Potential parameters: force constants for bonds (K_b, kcal/mol-Å^2), angles (K_a, in kcal/mod-rad^2), and torsions (K_t, in kcal/mol-rad^2); and ideal bond lengths and angles in Å and degrees.

K_b	b_0	
0.8	6.0	P–P along single strands
3.3	5.7	P–P along a strand within a helix
0.3	17.4	P–P across the helix, one base up
0.3	18.0	P–P directly across the helix
0.3	18.0	P–P across the helix, one base down
0.3	9.0	P–X inside a helix
5.0	3.8	Cα–Cα along a protein chain
K_a	θ_0	
7.45	100.0	P–P–P along strand, not in helix
7.45	151.7	P–P–P along strand within a helix
7.45	180.0	P–X–P within a helix
K_t	ϕ_0	
7.45	−107.7	P–P–P–P within a helix

listed in Table 1. They represent empirical potentials of mean force, based on the available experimental structural data [25]. Note that the harmonic torsional energy function generates pseudotorsions that are used to guarantee helix chirality.

Sequential bonds and angles are applied to the P-atoms along the RNA strand, and bonds are applied to all the base pairs (canonical or noncanonical). In helix stem regions, one torsion (e.g., i-j-m-n) and two additional bonds (e.g., i-m, j-n) are used for every two contiguous base pairs, and two bonds (e.g., i-q, q-n) and one angle (e.g., i-q-n) are used to connect the X-atom to its neighboring nucleotides (see Figure 1(b)).

The ribosomal proteins are treated essentially as rigid objects. They are represented by a modified version of an existing one-bead-per-residue model [30]. In our model, each amino acid residue is represented by a single pseudoatom (C-atom) centered on the α-carbon atom. Sequential bonds are applied to C-atoms along protein chains. In addition to its neighbor, each C-atom is connected to all other C-atoms in the protein that lie within a cutoff distance (8.0 Å) using harmonic pseudobonds:

$$K_s(r_{ij} - r_{ij}^o)^2. \tag{4}$$

The equilibrium bond lengths r_{ij}^o are obtained from the crystal structure of the 30S from *Thermus thermophilus* (1GIX) [15]. 5.0 kcal/mol/Å^2 is used for the force constant K_s based on normal mode analysis for the ribosomal proteins. This produces atomic root mean square fluctuations consistent with those derived from the crystallographic B-factors.

The semiharmonic potential is used for volume exclusions among the P-, X-, and C-atoms:

$$\begin{aligned} K_{nb}(r_{ij} - d_o)^2, &\quad r_{ij} < d_o, \\ 0, &\quad r_{ij} \geq d_o. \end{aligned} \tag{5}$$

This is simply a volume-exclusion term; specific interactions are modeled by distance-restraint terms discussed below. A force constant K_{nb} of 0.5 kcal/mol/A^2 is used for all pairs of atom types. The nonbond contact distances d_o are obtained by taking the minimal nonbond contact distances for every pair of atom types in the crystal structure. Values of d_o between two different atom types are shown in Table 2.

TABLE 2
Nonbond contact distances (Å).

	P	X	C
P	5.3	7.5	4.9
X		9.5	6.9
C			4.5

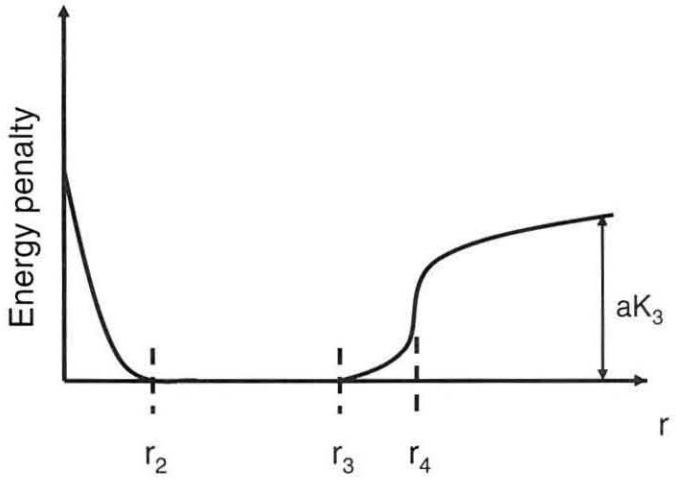


FIG. 2. *Energy penalty profile of distance restraints.*

2.2. Protein-RNA and helix-helix interactions. Morse-like distance restraints are used to represent specific protein-RNA and helix-helix interactions. The energy penalty profile is shown in Figure 2. r_2 and r_3 are the lower and upper bounds, respectively. When distance $r < r_2$, the energy E is parabolic with force constant K_2. When $r_2 \leq r \leq r_3$, $E = 0$. When $r_3 \leq r < r_4$, E is parabolic with force constant K_3. When $r \geq r_4$, E is hyperbolic given by

$$K_3 \left[\frac{b}{(r - r_3)} + a \right], \tag{6}$$

where $a = 3(r_4 - r_3)^2$ and $b = -2(r_4 - r_3)^3$. This function matches smoothly to the parabola at $r = r_3$ and tends to an asymptote of aK_3 at large r.

Such distance restraints can mimic the making and breaking of contacts in a dynamic molecular system. Since ribosomes are dynamic systems in which the protein-RNA and helix-helix contact distances are dynamically changing, the above distance restraints provide a rational way to represent specific protein-RNA and RNA-RNA interactions.

To set the protein-RNA distance restraints in the 30S subunit, we used the crystal structure of the 30S from *Thermus thermophilus* (1GIX) [15] and calculated all C-P pair distances for every protein and the RNA. Since proteins bind to the RNA mainly in the form of hydrogen bonds in which the corresponding C-P distance is about 9.0 Å, we pick up the C-P pairs with distance less than 10.0 Å, and the corresponding distance plus 0.5 Å is set as the upper bound r_3. The overall behavior of the number of restraints as the function of protein numbers (S2, S3, ..., S20) is consistent with the fraction of protein surface area packed against RNA obtained by the protein-RNA

surface area calculations for the 30S subunit [31].

There are many helix-helix interactions in the form of non-Watson–Crick hydrogen bonds in the crystal structure of the 16S RNA. They play important roles in forming the tertiary structure of the RNA. Wimberly et al. [14] listed all pairs of non-Watson–Crick hydrogen bonds of the 16S RNA based on a 3.0 Å-resolution X-ray crystal structure of the 30S from the *Thermus thermophilus* (PDB code: 1FJG). Those non-Watson–Crick hydrogen bonds inside helices are automatically included in our helix stem model. The other non-Watson–Crick hydrogen bonds (mainly among helices and inside loops) are considered to be distance restraints. The corresponding P-P distances are obtained from the 5.5 Å-resolution crystal structure (1GIX), and the distance plus 0.5 Å is set as the upper bound r_3.

We define $r_4 = r_3 + 0.75$ and $r_2 = r_3 - 2.0$. The force constants K_2 and K_3 are set 20 kcal/mol/Å^2. The force constants are optimized to maintain the 30S subunit around its crystal structure at room temperature, while allowing enough flexibility for the free 16S RNA.

The force field parameters shown in Table 1 are based on the general low-resolution force field for large RNAs and RNP particles developed by the Harvey lab [25]. The parameters were based on the statistical distributions of interphosphate distances, angles, and improper torsions in the crystal structure of phenylalanine tRNA. Analysis of later crystal structures (including the Group I ribozyme and the ribosome) has shown that the original parameters are still valid. This is not surprising, since the force field for the RNA component is basically designed to maintain the structure of the double-helical A-form RNA. Similarly, the protein parameters are generic, based on existing elastic network models. Only the protein-RNA interaction terms are specific to this system, and the force constants used there can take on a broad range of values. We found that setting K_2 and K_3 to values in the range of 5 to 20 kcal/mol/Å^2 produced nearly identical results; smaller force constants (such as 1 kcal/mol/Å^2) were too weak to enforce the correct protein-RNA interactions, whereas larger values were unreasonably stiff.

2.3. Force field assembly (FFA). FFA is one of the key modules in YUP. The FFA module defines methods to prepare and analyze a specific class of molecular model. It generates three objects, Potential, AtomMap, and AtomVector, by assembling and executing the appropriate fragments of code and drawing the appropriate parameters from a database file. The Potential and the AtomMap contain information essentially equivalent to those in the parameter/topology files in the traditional molecular modeling programs such as Amber and CHARMM. YUP generates several AtomVectors, but the only one required for Amber is the AtomVector containing all the coordinates. In YUP, FFA is still undergoing development, and new methods are being added to the module as needed. Using an early version of YUP, we completed FFA and the related modules for our specific low-resolution models for the ribosomal RNA and proteins. To connect YUP and Amber, a Python script (yup2amber.py in Figure 3) was developed to integrate reading in PDB files for an initial conformation, adding coordinates for X-atoms, running YUP modules for assembling force fields, and outputting Amber parameter and coordinate files for MD simulations. The program flow diagram is shown in Figure 3. (Note that the program flow in the recently released version of YUP is somewhat different from that shown in Figure 3.) In ReadyMade.py, molecules are defined in a hierarchical way based on the secondary structure. For example, the helix H3 in the 16S (see Figure 1(c)) is considered to be a subdomain which consist of two helix stems, H3stem1 and H3stem2, and one single

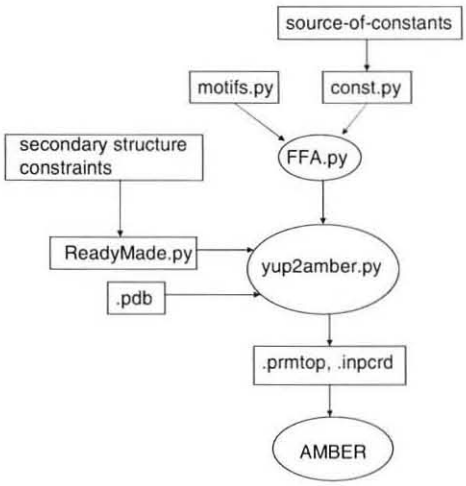


FIG. 3. *Program flow diagram.*

nucleotide. In H3stem1 and H3stem2, the first number corresponds to the starting P-atom of one strand of the helix stem, the third number corresponds to the starting P-atom of the other strand, and the second number is the number of base pairs or the length of the helix stem. In motifs.py, methods are defined for assembling force fields for each motif (TRACT, HELIX, DOMAIN, etc.) In const.py, each term in the potential functional is defined. Atomic properties and potential parameters are given in the source-of-constants database file.

2.4. MD simulations. The Amber 8 suite of programs [32] was used to perform all simulations and analyze trajectories. Langevin dynamics is used to mimic a solvent environment. A collision frequency of 2 ps^{-1} and a time step of 20 fs were used.

The starting structure of the 30S subunit of *Thermus thermophilus* was taken from the Protein Data Bank (PDB code: 1GIX) [15]. It is part of the crystal structure of the complete *Thermus thermophilus* 70S ribosome at 5.5 Å resolution. In our low-resolution model, the whole 30S subunit is represented by 4181 pseudoatoms (1519 P-atoms, 266 X-atoms, and 2396 C-atoms). The low-resolution tertiary structures from the 50S side and the solvent side are shown in Figure 4.

There is also a higher-resolution crystal structure of the 30S subunit (PDB code 1J5E); the root mean square deviation (RMSD) of the phosphate and protein Cα atoms between this and the structure we started with is 2.7 Å. This is smaller than the typical changes we are modeling here, so that it is unlikely that our results would be significantly affected by a change in initial structure.

We performed MD simulations for the bound and the unbound 16S RNA. For the bound 16S RNA, in addition to including all 30S proteins, we also performed simulations for the 16S RNA bounded by some selected proteins following the S7 assembly pathway. In each case, 10000 snapshots were extracted for analysis from the 100ns production run with an interval of 10 ps between snapshots.

All simulations were run on an SGI Altix 3700 Linux server. For the bound RNA (with all 30S proteins), it took about 44 hours to complete a 100ns simulation using 8 processors.

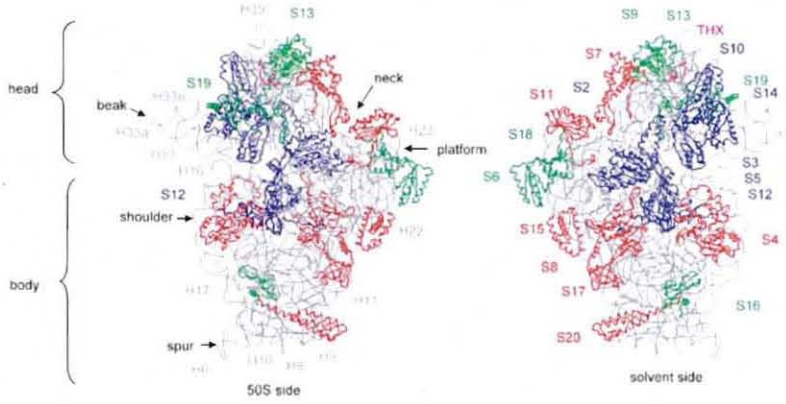


FIG. 4. *The crystal structure of the 30S subunit from Thermus thermophilus (PDB code: 1GIX). The 16S RNA is colored gray. Primary, secondary, and tertiary proteins are colored red, green, and blue, respectively. Protein THX is colored pink. Some helices are labeled in gray. The 5′ and 3′ minor domains form the body. The 3′ major domain forms the head. The central domain forms the platform. The spur is formed by helix H6. The shoulder is formed by helices H16 and H17. The beak is formed by helices H33, H33a, and H33b.*

3. Results and discussion.

3.1. Conformational changes of the 16S RNA and the roles of ribosomal proteins. Starting from the crystal structure (1GIX) (see Figure 4), we performed 100ns constant temperature (300K) MD simulations for the free 16S RNA and the RNA bound by 20 ribosomal proteins (S2, S3, ..., S20, THX). Snapshots of the conformations at 20ns, 40ns, 60ns, 80ns, and 100ns are shown in Figure 5. It can be seen that the conformation of the unbound 16S RNA greatly changes with the simulation time. It deviates from the crystal structure more and more. Long helices and loop regions show the most flexibilities. The initial tertiary structure is broken. As shown in Figure 6, at 100ns, the RMSD relative to the crystal structure is close to 40 Å. For the bound 16S RNA, however, the conformation is very stable. The overall shapes of all the snapshots are very similar to the initial crystal structure. As shown in Figure 6, the RMSD is always around 10 Å over the simulation time; this is then a rough estimate of the expected precision of our low-resolution models. Obviously, the ribosomal proteins are playing the role of maintaining the 16S RNA in its native structure in the ribosome. As seen in Figure 5, those regions which have more contacts with the proteins are more stable. Those regions of the RNA that have few or no contacts with proteins are more flexible and deviate more from the initial structure. For example, helices H6, H16, H17, H33, H33a, H33b, and H44 have very few contacts with proteins, so they have higher flexibilities than other regions of the RNA.

Figure 7 shows the RMSD contributions from each domain of the bound RNA. The central domain is the most stable one and has the smallest average RMSD ($\sim$ 7.5 Å). The 3′ minor domain is the most flexible one and has the largest average RMSD ($\sim$ 12 Å). Most of the time, the 5′ domain is very stable and has a small RMSD as

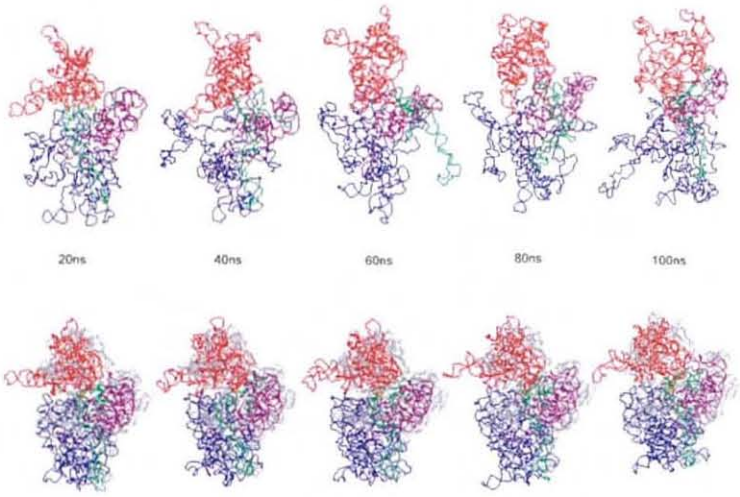


FIG. 5. *Snapshots at* 20*ns*, 40*ns*, 60*ns*, 80*ns*, *and* 100*ns for the free* 16*S RNA (top) and the full RNA/protein complex of the* 30*S subunit (bottom) with RNA domains colored as in the secondary structure in Figure* 1*(a). Proteins are colored gray.*

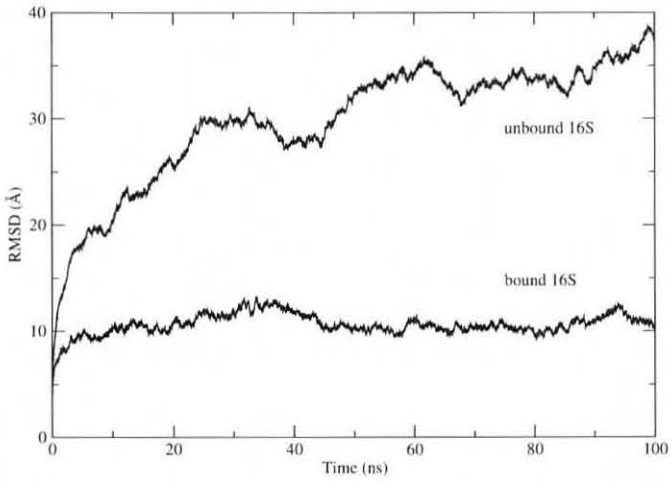


FIG. 6. *RMSDs for the bound and the unbound* 16*S RNA. (All RMSDs are relative to the initial crystal structure (*1*GIX) in Figure* 4.*)*

the central domain, but it transiently has a very large RMSD as the 3′ minor domain. The 3′ major domain is relatively stable with the RMSD between the values of the central domain and the 3′ minor domain.

The 30S proteins are usually classified into three groups [33]. Those proteins which can directly bind to the 16 RNA are called primary proteins. Those that can bind to the RNA after at least one primary protein bind are called secondary proteins. Tertiary proteins are those that can bind to the RNA after at least one primary

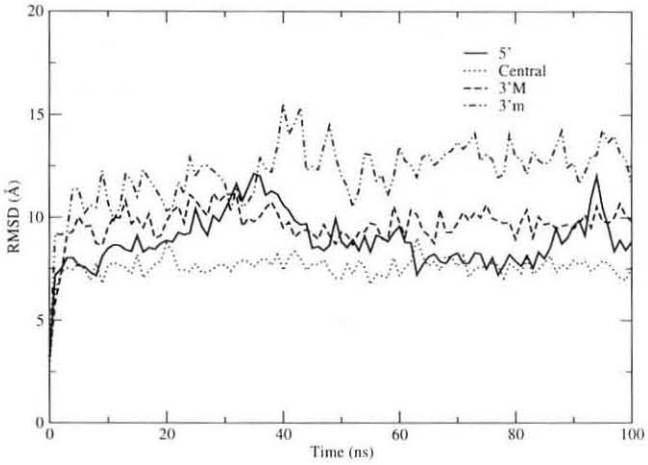


FIG. 7. *RMSDs for 5′, central, 3′ major, and 3′ minor domains of the bound 16S RNA.*

protein and one secondary protein are bound. For the *Thermus thermophilus* 30S, primary proteins include S4, S7, S8, S11, S15, S17, S20; secondary proteins include S6, S9, S13, S16, S18, S19; tertiary proteins include S2, S3, S5, S10, S12, S14. Studies have shown that primary proteins play the most important role in the ribosomal assembly and secondary proteins play a more important role than tertiary proteins [17, 33, 34, 35].

The central domain is smaller than the 5′ and 3′ major domains and has a compact conformation. Five proteins closely and uniformly bind to the central domain, including three primary proteins (S15, S8, S11) and two secondary proteins (S6 and S18). Those particular protein binding and conformational properties of the central domain make it the most stable domain in the 30S subunit. The 5′ domain has six proteins, including three primary proteins (S20, S17, S4). Thus most of the time it has RMSD behavior similar to the central domain. But the 5′ domain has larger size than the central domain, and its lower part is bound only by two proteins (one primary S20 and one secondary S16). Also, it has the extended spur and shoulder parts. Those properties make the 5′ domain more flexible than the central domain. The 3′ major domain has about the same size as the 5′ domain. Although eight proteins bind to it, most of them are tertiary and secondary proteins, and there is only one primary protein (S7). In addition, it has an extended beak. Thus it has a larger RMSD than the central and 5′ domains most of the time. The 3′ minor domain is mainly formed by the long helix H44, which is the longest single helix in the 16S RNA. However, H44 has only few contacts with proteins S20, S5, and S12. Thus it is the most flexible domain.

The central domain has been analyzed in detail both experimentally and theoretically. Orr, Hagerman, and Williamson [36] studied the conformational changes of the 16S RNA from the *Bacillus stearothermophilus* induced by the binding of the ribosomal protein S15 and Mg^{2+}. They measured the interhelical angles of helices H20, H21, and H22 and found that the free junction of helices H20, H21, and H22 is planar with about 120° interhelical angles, whereas S15 and Mg^{2+} binding induces a junction conformation in which two helices, H21 and H22, become colinear and the third, helix H20, forms a 60° angle with respect to helix H22. By means of all-atom MD

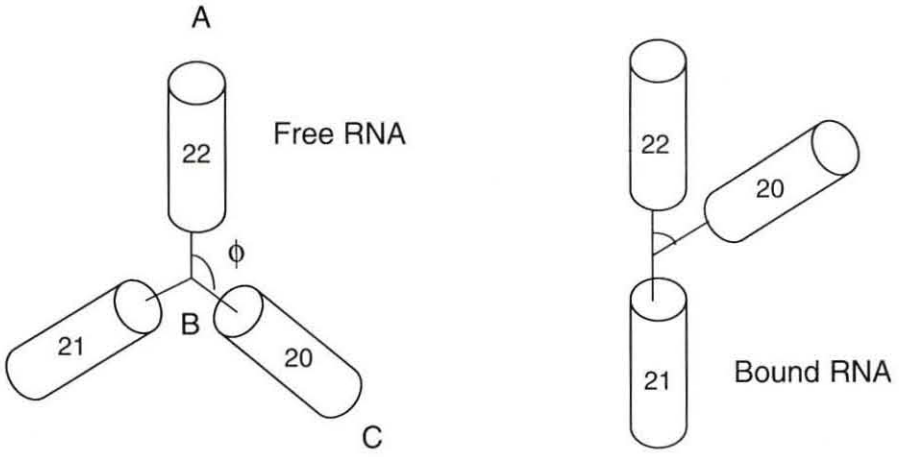


FIG. 8. *Schematic representation of the relative orientations of helices H20, H21, and H22. Angle ϕ is defined by three mass centers of nucleotide P-atoms: A(A663,G742), B(A753,C754,G587), C(G758,A759,U582,A583). Nucleotide numbering is the same as in Figure 1(a).*

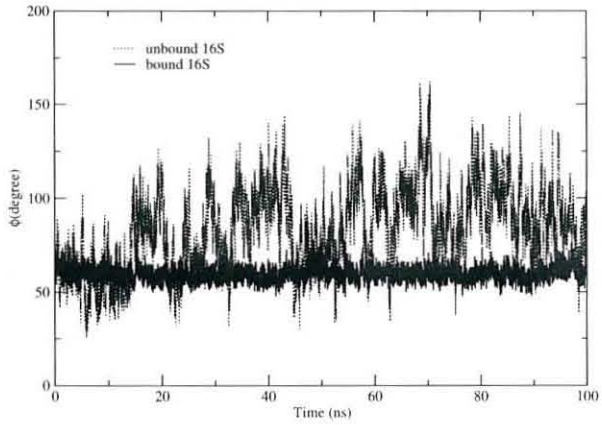


FIG. 9. *Time dependence of angle ϕ between helices H20 and H22 for the bound and the unbound 16S RNA.*

simulations, Li, Ma, and Shapiro [18] studied the conformational changes of the core central domain of the 16S RNA from *Thermus thermophilus* induced by the binding of S15 protein. They calculated the interhelical angle between helices H20 and H22 and the width of the major groove of helix H22 as the function of simulation time for the free and the bound central domain. They found that in the absence of S15, the interhelical angle is increased from 80° in the crystal structure to 114° in the average structure of the unbound RNA, and the width of the major groove greatly fluctuates and most of the time it is much wider than in the crystal structure.

Our low-resolution MD simulations also predict conformational changes consistent with the above experimental studies and high-resolution MD simulations. We used the definition of [18] for the interhelical angle between helices H20 and H22 (see Figure 8) and the angle as the function of simulation time for the free and the bound 16S RNA is shown in Figure 9. It can be seen that, for the bound 16S RNA, the angle is

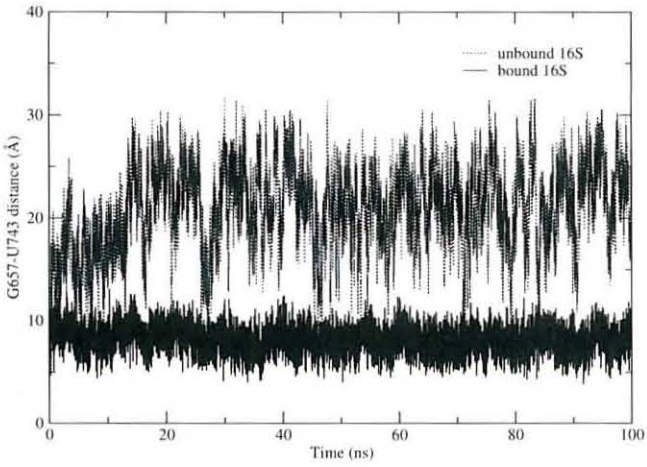


FIG. 10. *Time dependence of the distance between G657 :P and U743 :P. Nucleotide numbering is the same as in Figure 1(a).*

stabilized around the value corresponding to that in the crystal structure (60°), while, for the unbound 16S RNA, the angle greatly fluctuates. The angle of the unbound 16S RNA may transiently adopt the value of the crystal structure, but, most of the time, it adopts much larger values up to 120°.

Helix H22 consists of three helix stems and two bulge loops. In our model, each helix stem is essentially rigid, but the entire helix is not necessarily rigid. The width of the "major groove" of helix H22 (around the bulge) can be defined as the distance between two P-atoms in the two nucleotides G657 and U743 [18]. From Figure 10, one can see that, for the bound 16S RNA, the width is small and stabilized around 9.0 Å, close to the distance in the crystal structure, while, for the unbound 16S RNA, the width fluctuates greatly, and most of the time it is much wider than that of the bound 16S RNA.

In summary, our MD simulations indicate that the ribosomal proteins play very important roles in maintaining the 16S RNA around its crystal structure. The bound 16S RNA always adopts its crystal-like conformation. The more protein contacts that a given RNA region has, the more stable the region. Primary proteins help more in stabilizing the conformation of the RNA than other proteins. Without the ribosomal proteins, the 16S RNA tends to adopt extended conformations which greatly deviate from its crystal structure. However, some local domains may transiently adopt the crystal-like conformations.

3.2. The S7 assembly pathway. In the early 1970s, Nomura and coworkers studied the sequence of protein binding in the 30S ribosomal subunit of *Escherichia coli* with *in vitro* reconstitution [33]. These studies have been extended recently with modern analysis methods, including mass spectroscopy and more classical biochemical separation techniques [35, 37, 38]. It has generally been found that the proteins bind to the 16S RNA in a sequential and cooperative way, roughly as shown in the Nomura assembly map (see Figure 11). From the map, one can see that primary proteins S4, S7, S8, S15, S17, and S20 first independently bind to the 16S RNA, followed by secondary proteins S6, S9, S16, S18, and S19 and tertiary proteins S2, S3, S5, S10, S11, S12, S13, S14, and S21. Also one can see three relatively independent assembly

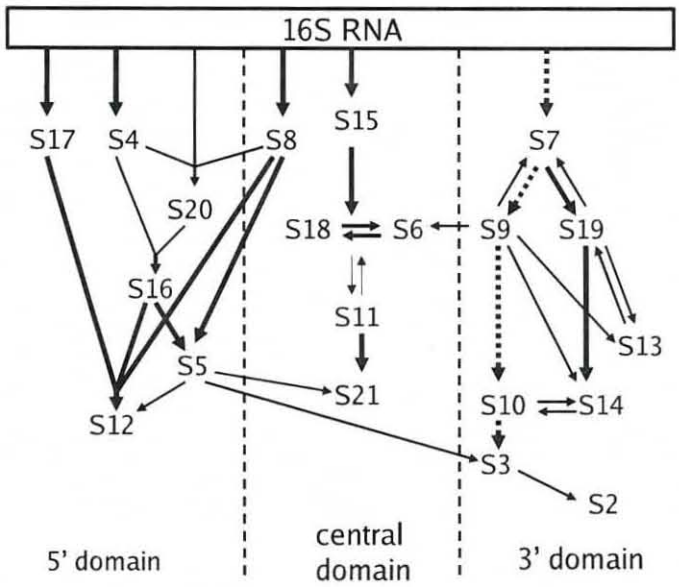
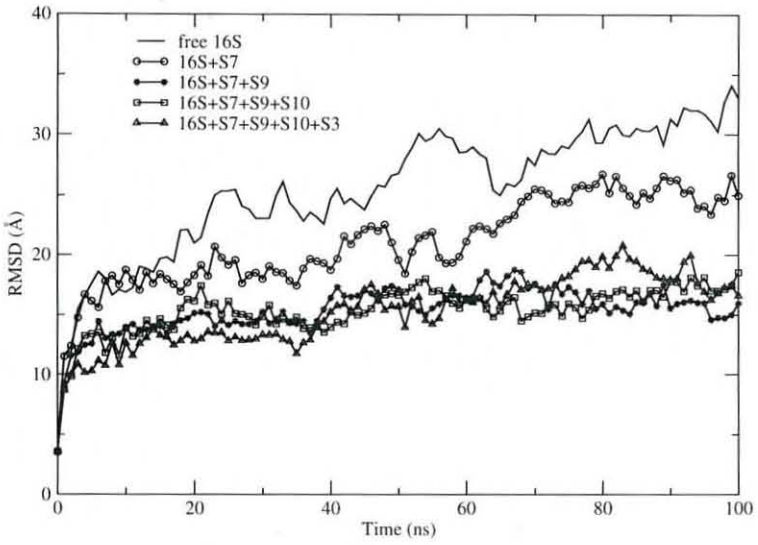


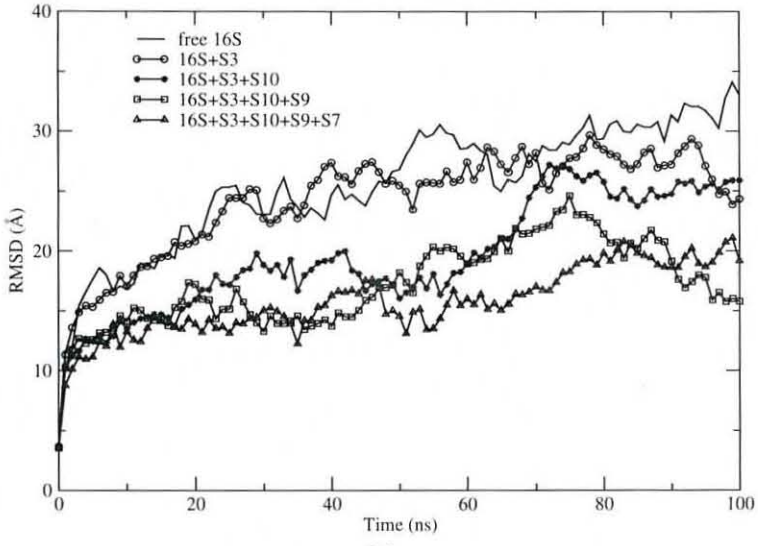
FIG. 11. *Nomura assembly map for Escherichia coli. This map was drawn based on the original map in* [33] *and the map in* [35]. *It reflects current knowledge of the location of 30S proteins in the subunit and the order of domain assembly (from the 5′ to the 3′ domain). Arrows indicate protein binding dependencies. The thickness of the arrows represents the strength of the dependencies. The pathway indicated by dashed arrows is analyzed in this work.*

pathways: (S17, S4, S20, S8), S15, and S7. The binding sites of the proteins in the pathways are located in the 5′, central, and 3′ domains of the 16S RNA, respectively. In order to understand why protein binding follows the Nomura map, we examined the S7 pathway by comparing the conformational changes of the 16S RNA when proteins bind to the RNA in different orders. Since proteins on the S7 pathway mainly bind to the 3′ major domain, we will compare the RMSD changes of the 3′ major domain bound by proteins in the order of S7, S9, S10, and S3 and in the reverse order.

First, we ran a simulation with only protein S7 allowed to bind to the RNA. As seen in Figure 12(a), in comparison with the free RNA, the bound 3′ major domain has a smaller RMSD most of the time. Then we carried out a simulation allowing proteins S7 and S9 to bind to the RNA. The corresponding RMSD is remarkably decreased. Next, in the simulation of S7, S9, and S10 binding to the RNA, the RMSD behavior is almost the same as that with just S7 and S9 binding. Finally, when S7, S9, S10, and S3 bind to the RNA, again there is no overall improvement in the RMSD fluctuations. The results clearly indicate that the primary and secondary proteins S7 and S9 play more important roles in stabilizing the conformation of the 3′ major domain than the tertiary proteins S10 and S3. The binding of S7 and S9 not only constrains their contact regions in the RNA but also helps to organize the binding sites for later binding proteins S10 and S3. Therefore, protein binding in this order would be more efficient for the 30S assembly. This is consistent with the results of the Monte Carlo simulations of Stagg, Mears, and Harvey [17]. These results can be confirmed by allowing the proteins to bind in the reverse order. As seen in Figure 12(b), if we allow S3 binding first, the overall RMSD is almost the same as that of the free 16S RNA. After S10 binding, the overall RMSD is reduced somewhat during



(a)



(b)

FIG. 12. *RMSDs of the 3′ major domain for free 16S RNA and bound 16S RNA with proteins on the S7 pathway.* (a) *Proteins bind to RNA following the order of the S7 pathway.* (b) *Proteins bind to RNA in the reverse order of the S7 pathway.*

parts of the simulation, though not through all of it. Only after S7 and S9 binding do we obtain significantly reduced RMSD fluctuations, confirming the importance of the correct binding order.

Our results suggest structural rationales for ribosomal proteins to bind to the RNA in a sequential way. The binding orders in the Nomura map appear to facilitate 30S assembly because primary and secondary proteins not only constrain their contact regions in the RNA but also help to organize the binding sites for tertiary pro-

teins. More extensive studies of other parts of the assembly pathway will be reported elsewhere.

4. Concluding remarks. We have developed methodology and tools for performing very long MD simulations of large biological assemblies such as ribosomes at a low-resolution level. Our procedure combines the coarse-grained modeling capabilities of YUP with the parallelized MD code in Amber, allowing about 60 nsec/day of simulation to be carried out on 8 CPUs. As our first application, we studied structural and dynamical properties of the 30S subunit from the *Thermus thermophilus*. We found that ribosomal proteins play critical roles in maintaining the 16S RNA around its crystal structure. Primary proteins help more in stabilizing the conformation of the RNA than secondary and tertiary proteins. Protein binding orders in the Nomura map are favorable for 30S assembly because primary and secondary proteins not only constrain their contact regions in the RNA but also help to organize the binding sites for tertiary proteins. These low-resolution MD simulations produce results that are qualitatively consistent with those from experimental and from other simulations. Our methodology and tools provide efficient ways for exploring structural and dynamical questions in large macromolecular assemblies with MD simulations.

Acknowledgments. We thank Scott M. Stagg for providing a Python script for reading PDB files and for helpful discussions, Francois Dupradeau for advice on using the Insight program, and Joanna Trylska and Jamie Williamson for helpful discussions.

REFERENCES

[1] A. Expert-Bezançon and P. L. Wollenzien, *Three-dimensional arrangement of the Escherichia coli 16S ribosomal RNA*, J. Mol. Biol., 184 (1985), pp. 53–66.

[2] K. Nagano, M. Harel, and M. Takezawa, *Prediction of three-dimensional structure of Escherichia coli ribosomal RNA*, J. Theoret. Biol., 134 (1988), pp. 199–256.

[3] R. Brimacombe, *The emerging three-dimensional structure and function of 16S ribosomal RNA*, Biochemistry, 27 (1988), pp. 4207–4214.

[4] S. Stern, B. Weiser, and H. F. Noller, *Model for the three-dimensional folding of 16S ribosomal RNA*, J. Mol. Biol., 204 (1988), pp. 447–481.

[5] M. I. Oakes, L. Kahan, and J. A. Lake, *DNA-hybridization electron microscopy tertiary structure of 16S rRNA*, J. Mol. Biol., 211 (1990), pp. 907–918.

[6] H. F. Noller, R. Green, G. Heilek, V. Hoffarth, A. Huttenhofer, S. Joseph, I. Lee, K. Lieberman, and A. Mankin, C. Merryman, T. Powers, E. V. Puglisi, R. R. Samaha, and B. Weiser, *Structure and function of ribosomal RNA*, Biochem. Cell Biol., 73 (1995), pp. 997–1009.

[7] F. Mueller, T. Doring, T. Erdemir, B. Greuner, N. Junke, M. Osswald, J. Rinke-Appel, K. Stade, S. Thamm, and R. Brimacombe, *Getting closer to an understanding of the three-dimensional structure of ribosomal RNA*, Biochem. Cell Biol., 73 (1995), pp. 767–773.

[8] A. Malhotra, R. K.-Z. Tan, and S. C. Harvey, *Prediction of the three-dimensional structure of Escherichi coli 30S ribosomal subunit: A molecular mechanics approach*, Proc. Natl. Acad. Sci. USA, 87 (1990), pp. 1950–1954.

[9] T. R. Easterwood and S. C. Harvey, *Modeling the structure of the ribosome*, Biochem. Cell Biol., 73 (1995), pp. 751–756.

[10] A. Malhotra and S. C. Harvey, *A quantitative model of the Escherichia coli 16S RNA in the 30S ribosomal subunit*, J. Mol. Biol., 240 (1994), pp. 308–340.

[11] J. H. Cate, M. M. Yusupov, G. Z. Yusupova, T. N. Earnest, and H. F. Noller, *X-ray crystal structures of 70S ribosome functional complexes*, Science, 285 (1999), pp. 2095–2104.

[12] F. Schluenzen, A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi, and A. Yonath, *Structure of functionally activated small ribosomal subunit at 3.3 A resolution*, Cell, 102 (2000), pp. 615–623.

[13] N. Ban, P. Nissen, J. Hansen, P. B. Moore, and T. A. Steitz, *The complete atomic structure of the large ribosomal subunit at* 2.4 *Å resolution*, Science, 289 (2000), pp. 905–920.

[14] B. T. Wimberly, D. E. Brodersen, W. M. Clemons, Jr., R. J. Morgan-Warren, A. P. Carter, C. Vonrhein, T. Hartsch, and V. Ramakrishnan, *Structure of the 30S ribosomal subunit*, Nature, 407 (2000), pp. 327–339.

[15] M. M. Yusupov, G. Zh. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. D. Cate, and H. F. Noller, *Crystal structure of the ribosome at* 5.5 *Å resolution*, Science, 292 (2001), pp. 883–896.

[16] B. S. Schuwirth, M. A. Borovinskaya, C. W. Hau, W. Zhang, A. Vila-Sanjurjo, J. M. Holton, and J. H. D. Cate, *Structures of the bacterial ribosome at* 3.5 *Å resolution*, Science, 310 (2005), pp. 827–834.

[17] S. M. Stagg, J. A. Mears, and S. C. Harvey, *A structural model for the assembly of the 30S subunit of the ribosome*, J. Mol. Biol., 328 (2003), pp. 49–61.

[18] W. Li, B. Ma, and B. A. Shapiro, *Binding interactions between the core central domain of 16S rRNA and the ribosomal protein S15 determined by molecular dynamics simulations*, Nucleic Acid Res., 31 (2003), pp. 629–638.

[19] F. Tama, M. Valle, J. Frank, C. L. Brooks, III, *Dynamic reorganization of the functionally active ribosome explired by normal code analysis and cryo-electron microscopy*, Proc. Natl. Acad. Sci. USA, 100 (2003), pp. 9319–9323.

[20] Y. Wang, A. J. Rader, I. Bahar, and R. L. Jernigan, *Global ribosome motions revealed with elastic network model*, J. Struct. Biol., 147 (2004), pp. 302–314.

[21] J. Trylska, V. Tozzini, J. A. McCammon, *Exploring global motions and correlations in the ribosome*, Biophys. J., 89 (2005), pp. 1455–1463.

[22] W. Smith and T. Forester, *DL-POLY-2.0: A general-purpose parallel molecular dynamics simulation package*, J. Mol. Graphics, 14 (1996), pp. 136–141.

[23] K. Y. Sanbonmatsu, S. Joseph, and C.-S. Tung, *Simulating movement of tRNA into the ribosome during decoding*, Proc. Natl. Acad. Sci. USA, 102 (2005), pp. 15854–15859.

[24] R. K.-Z. Tan and S. C. Harvey, *Yammp: Development of a molecular mechanics program using the modular programming method*, J. Comput. Chem., 14 (1993), pp. 455–470.

[25] A. Malhotra, R. K.-Z. Tan, and S. C. Harvey, *Modeling large RNAs and ribonucleoprotein particles using molecular mechanics techniques*, Biophys. J., 66 (1994), pp. 1777–1795.

[26] http://rumour.biology.gatech.edu/YammpWeb/.

[27] R. K.-Z. Tan, S. Petrov, and S. C. Harvey, *YUP: A molecular simulation program for coarse-grained and multiscaled models*, J. Chem. Theory Comput., 2 (2006), pp. 529–540.

[28] http://www.rna.icmb.utexas.edu.

[29] R. R. Gutell, *Collection of small subunit (16S and 16S-like) ribosomal RNA structures:* 1994, Nucleic Acid Res., 22 (1994), pp. 3502–3507.

[30] T. Haliloglu and I. Bahar, *Coarse-grained simulations of conformational dynamics of proteins: Application to Apomyglobin*, Proteins, 31 (1998), pp. 271–281.

[31] D. E. Brodersen, W. M. Clemons, Jr., A. P. Carter, B. T. Wimberly, and V. Ramakrishnan, *Crystal structure of the 30S ribosomal subunit from Thermus thermophilus: Structure of the proteins and their interactions with 16S RNA*, J. Mol. Biol., 316 (2002), pp. 725–768.

[32] D. A. Case, T. E. Cheatham, III, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, and R. Woods, *The Amber biomolecular simulation programs*, J. Comput. Chem., 26 (2005), pp. 1668–1688.

[33] W. A. Held, B. Ballou, S. Mizushima, and M. Nomura, *Assembly mapping of 30S ribosomal proteins from Escherichia coli*, J. Biol. Chem., 249 (1974), pp. 3103–3111.

[34] S. C. Agalarov, G. S. Prasad, P. M. Funke, C. D. Stout, and J. R. Williamson, *Structure of the S15, S6, S18-rRNA complex: Assembly of the 30S ribosome central domain*, Science, 288 (2000), pp. 107–112.

[35] M. W. T. Talkington, G. Siuzdak, and J. R. Williamson, *Assembly landscape of the 30S ribosomal subunit*, Nature, 438 (2005), pp. 628–632.

[36] J. W. Orr, P. J. Hagerman, and J. R. Williamson, *Protein and Mg(2+)-induced conformational changes in the S15 binding site of 16S ribosomal RNA*, J. Mol. Biol., 275 (1998), pp. 453–464.

[37] G. M. Culver, *Assembly of the 30S ribosomal subunit*, Biopolymers, 68 (2003), pp. 234–249.

[38] K. L. Holmes and G. M. Culver, *Analysis of conformational changes in 16S rRNA during the course of 30S subunit assembly*, J. Mol. Biol., 354 (2005), pp. 340–357.